

Gutachten

Version: 1.0 (Release)
Date: 13.10.2010

„Analyse eines Fernsehmitschnitts“ (Angebotsnummer: A82/10/0052)

Fraunhofer Institut für Digitale Medientechnologie

Ehrenbergstraße 31

D-98693 Ilmenau



1 Ausgangslage

Begutachtet wird ein Audioausschnitt von ca. 24 Sekunden Länge aus vom ORF veröffentlichtem Filmmaterial¹. Im Allgemeinen soll bewertet werden, ob es sich beim vorliegenden Material um unbearbeitetes Originalmaterial handelt. Die Betrachtungen konzentrieren sich dabei auf die Audiospur des Ausschnitts, welche im Folgenden als **Rohmaterial** bezeichnet wird.

Neben dem vorliegenden Ausschnitt wurde auch eine Aufnahme des vom ORF ausgestrahlten Beitrags in der Sendung „Schauplatz“ herangezogen, welche im Folgenden als **Sendungsmitschnitt** bezeichnet wird.

Das Material zeichnet sich dadurch aus, dass eine Dialogähnliche Situation zwischen den Beteiligten Personen vorliegt. Die Sätze und Worte sind dabei vergleichsweise stark mit Nebengeräuschen überlagert, da eine starke Hintergrundgeräuschkulisse durch Unterhaltungen nebenstehender Personen erzeugt wird.

2 Audioqualität

Das komplette Audiomaterial des Sendungsmitschnitts als auch des Rohmaterials liegt den Gutachtern im AIFF-Format² vor. Es lässt sich jedoch an Codierungsartefakten erkennen, dass es ursprünglich mit Hilfe eines verlustbehafteten Audiocodecs abgespeichert wurde. Dies wird auch durch die Metadaten der korrespondierenden Videodateien bestätigt, dort liegen Audio und Video gemultiplext als MP4³-Datei vor. Die Audiospuren sind jeweils mittels AAC LC⁴ codiert. Die Audiospuren entsprechen demnach nicht CD-Qualität.

3 Subjektive Begutachtung

Bei der Analyse des Materials, fällt auf, dass sich die Perspektive der Kamera deutlich von der Audio-Perspektive unterscheidet. Offensichtlich wurde der Ton von verschiedenen, nicht an der Kamera befestigten Mikrofonen aufgenommen. Insbesondere die Personen, die im Laufe des Ausschnitts angesprochen werden müssen eindeutig mit Ansteckmikrofonen ausgestattet zu sein. Das manifestiert sich darin, dass sie sich zeitweise von der Kamera wegbewegen, während man noch immer recht deutlich ihre Stimmen vernehmen kann. Dieser Eindruck wird bestätigt,

¹ <http://kundendienst.orf.at/unternehmen/news/amschauplatz.html> (abgerufen am 5. Oktober 2010)

² http://swlab.et.fh-duesseldorf.de/pc_pool/lernmodule/multimediateien/Kapitel31.htm (abgerufen am 07. Oktober 2010)

³ <http://de.wikipedia.org/wiki/MPEG-4> (abgerufen am 07. Oktober 2010)

⁴ http://www.iis.fraunhofer.de/fhg/Images/MPEG_AAC_Family_V0609_270809_EN_tcm97-67331.pdf (abgerufen am 07. Oktober 2010)

wenn man das vorliegende Rohmaterial mit dem ausgestrahlten Sendungsmitschnitt vergleicht:

Timecode (Ausschnitt)	Rohmaterial	Sendungsmitschnitt
(-00:00:01:00 bis 00:00:00:00)	Ausatmen kaum zu hören	Ausatmen deutlich zu hören
00:00:00:00 bis 00:00:04:00	Frage des Journalisten „und was sagst du zu ihm?“ deutlich zu hören	Frage weniger deutlich zu hören. Dafür sehr direkter Ton des Antwortenden.

Im Verlauf des Ausschnitts haben die Gutachter den Eindruck, dass sich im Rohmaterial die Mischungsverhältnisse der jeweiligen Mikrofone zueinander einige Male ändern.

Timecode (Ausschnitt)	Rohmaterial
00:00:03:00 bis 00:00:06:00	Ausblendung eines Mikrofons mit betonteren hochfrequenten (brillianten) Anteilen
00:00:18:00 bis 00:00:21:00	Einblendung eines Mikrofons mit betonteren hochfrequenten (brillianten) Anteilen

Des Weiteren kommen im Rohmaterial einige Geräusche oder Laute vor, die relativ auffällig sind. Einige dieser Geräusche passieren zeitlich gesehen innerhalb des Abschnitts, zu dem aufgrund eines Umschnitts kein Vergleichsmaterial in dem Sendungsmitschnitt vorliegt.

Timecode (Ausschnitt)	Rohmaterial
00:00:10:00 bis 00:00:13:00	Das Ausatmen einer der Personen ist recht deutlich zu vernehmen. Derartige Geräusche tauchen auch an anderen Stellen der Aufzeichnung auf. Der ungewöhnliche Klangeindruck rührt daher, dass in unmittelbarer Nähe eines Mikrofons ausgeatmet wurde.
00:00:13:00 bis 00:00:14:00	Ein hoher Piepston stammt von einer Digitalkamera oder einem Foto-Handy. Kurze Zeit später ist auch ein typisches synthetisches Kameralinsengeräusch zu hören, das aber teilweise von einer Aussage einer Person überdeckt wird.
00:00:20:00 bis 00:00:21:00	Im Hintergrund ist sehr leise und entfernt eine Aussage zu hören, deren Wortlaut nicht zu vernehmen ist, da er auch von der Stimme einer Person im Vordergrund überdeckt wird.

4 Zeit-Frequenz Analyse

Zum Zwecke der technischen Analyse wurden die beiden Aufnahmen mit Hilfe der Software Adobe Audition manuell synchronisiert, d.h. die beiden Audioaufnahmen wurden auf einer Zeitachse relativ zueinander so angeordnet, dass sich größtmögliche Übereinstimmung zwischen den beiden Schallwellenformen ergibt. Dafür mussten in den Sendungsmitschnitt ab Sekunde Sekunde 7,2 ca. 7,8 Sekunden Stille eingefügt werden. An dieser Stelle erfolgt im Sendungsmitschnitt ein Umschnitt, d.h. es liegt keine Audiomaterial vor, weswegen hier Stille eingefügt werden musste um bei den nachfolgenden Wellenformen der beiden Aufnahmen wieder einen synchronisierten Zustand zu erlangen.

4.1 Zeitsignalanalyse

In Abbildung 1 wird zum Vergleich die Wellenform des Rohmaterials der Wellenform des Sendungsmitschnitts gegenüber gestellt. Die eingefügte Stille ist deutlich erkennbar. Auch ist erkennbar, dass die Dynamik im Sendungsmitschnitt deutlich komprimiert wurde. Das bedeutet es wurde mittels eines Kompressors für einen weniger stark schwankenden Pegel gesorgt. Der Einsatz eines Kompressors ist in der Medienproduktion durchaus üblich. An auffälligen Ausschlägen der Signalhüllkurve liegt zumeist eine Wortäußerung vor, die durch entsprechende Textfelder annotiert wurde.

4.2 Annotation im Zeit-Frequenz Bereich

Zur Nachvollziehbarkeit der im Abschnitt 3 getroffenen Aussagen wird in Abbildung 2 eine Zeit-Frequenzdarstellung des Rohmaterials dargestellt. Die hörbaren Satzketten, Worte und Geräusche sind grafisch markiert und annotiert. Bereits in dieser Abbildung fällt auf, dass die hochfrequenten Anteile in der Mitte des Ausschnittes geringere Amplitude aufweisen als zu Beginn und Anfang. Das untermauert die Feststellung aus der subjektiven Analyse, dass ein eine Mikrofonspur mit brillanterem Klang aus- und wieder eingeblendet wird.

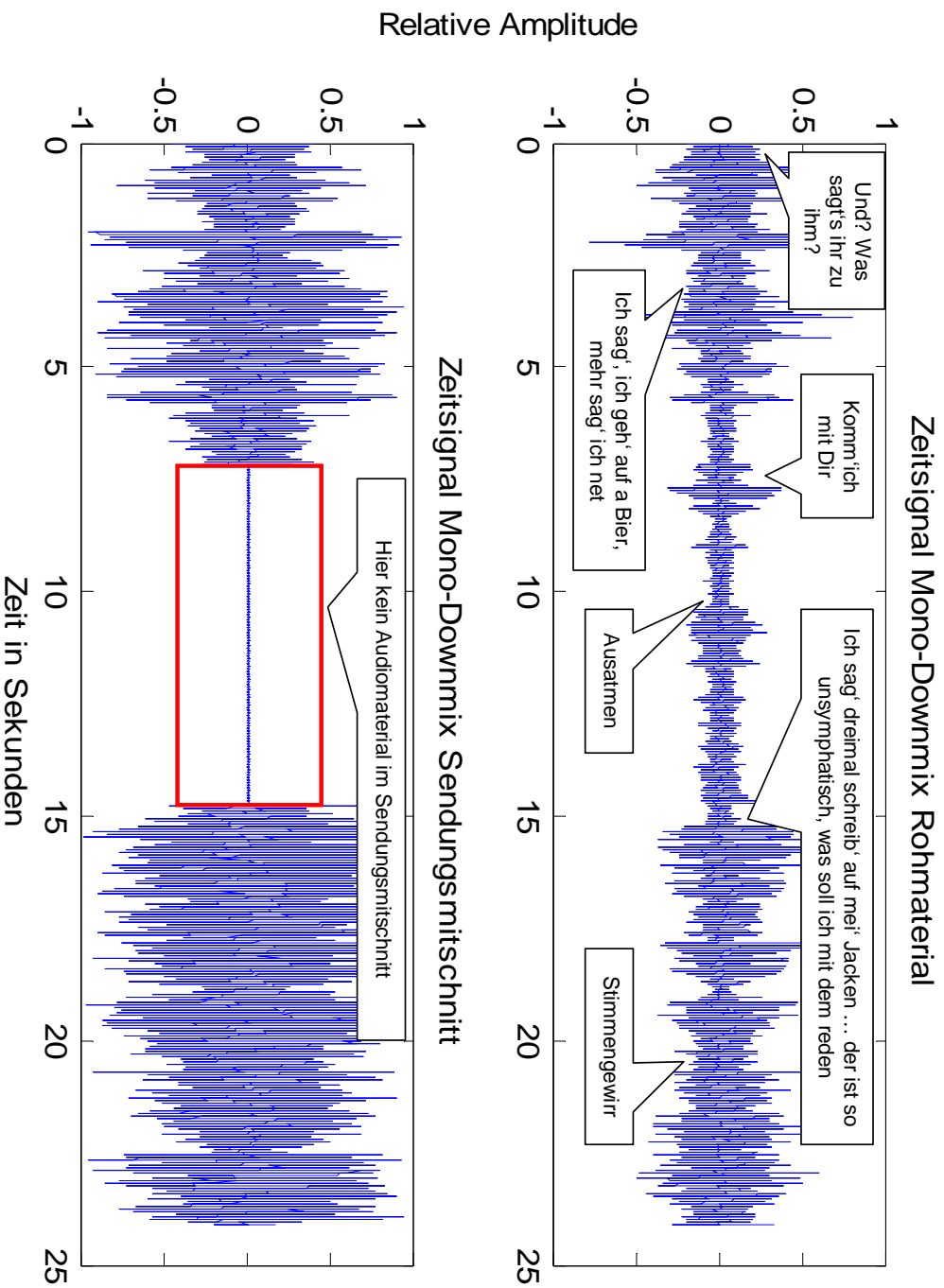


Abbildung 1: Vergleich der Zeitsignale Rohmaterial und Sendungsmitschnitt

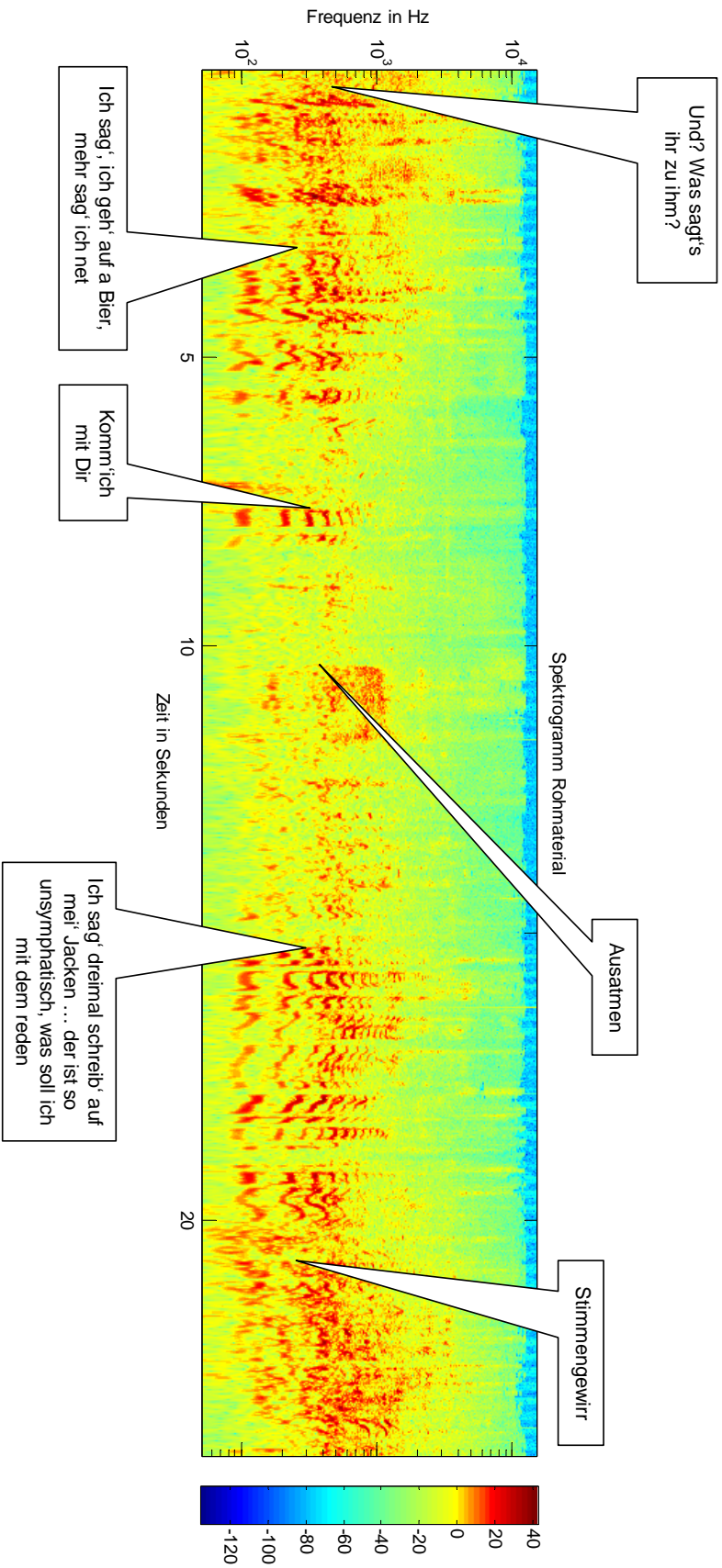


Abbildung 2: Annotiertes Spektrogramm des Rohmaterials

4.3 Entmischung der Mikrofonsignale

Da Rohmaterial und Sendungsmitschnitt zeitlich synchronisiert vorliegen können die offensichtlichen Pegelunterschiede bei der Abmischung der beiden Audiosignale ausgenutzt werden um den Signalanteil des Mikrofons mit den betonteren hochfrequenten Anteilen zu separieren (im Folgenden als Mikrofon 2 bezeichnet). Durch eine Non-Negative Matrix Factorization⁵ im Spektralbereich wird es möglich, diesen Signalanteil deutlicher hörbar zu machen. Es soll darauf hingewiesen werden, dass diese Separationsmethode nur für die ersten 7 Sekunden des Rohmaterials durchgeführt werden kann, da ab dann kein Vergleichsmaterial aus dem Sendungsmitschnitt vorliegt.

In diesem separierten Signal von Mikrofon 2 ist auch die Frage des Journalisten deutlicher zu hören. In Abbildung 3 werden der zeitliche Verlauf der Amplitudenhüllkurven der beiden Signale Mikrofon 1 (deutlich hörbare Aussagen der befragten Personen) und Mikrofon 2 (deutlich hörbare Frage des Journalisten) dargestellt. Zusätzlich wurde in rot auch jeweils eine durch lineare Regression ermittelte Gerade eingezeichnet, die den Trend des Hüllkurvenverlaufes wiedergibt. Mikrofon 2 ist dabei eindeutig bis zum Nullpegel fallend, während Mikrofon 1 einigermaßen konstanten Pegel aufweist. Dieses Messergebnis bestätigt eindeutig die subjektive Analyse, dass hier während der ersten 7 Sekunden die brillanter klingende Audiospur ausgeblendet wird.

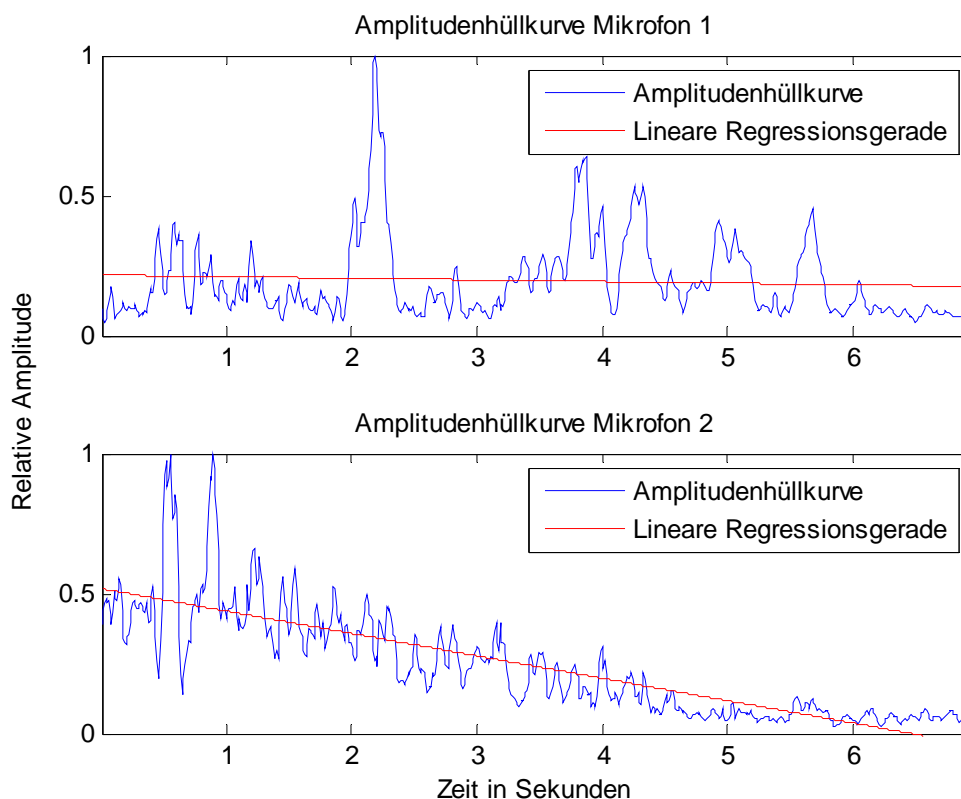


Abbildung 3: Vergleich der Amplitudenhüllkurven

⁵ Virtanen, T.: "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria". IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 3, pp. 1066–1074, 2007.

4.4 Störgeräuschunterdrückung

Da insbesondere das Geräusch des Ausatmens in der Mitte des Rohmaterials recht ungewöhnlich klingt wurde es eingehender untersucht. Mittels spektraler Subtraktion⁶ lassen sich die quasi-statischen Spektralanteile des Ausatmens vom Rest des Gesamtklangs abziehen. Dies wurde mit zwei verschiedenen Stärkegraden der Reduktion durchgeführt. In Abbildung 4 sind die Spektrogramme der beiden resultierenden Signale dargestellt. Es ist ersichtlich dass durch die Bearbeitung viele Artefakte (so genannte Musical Noise) entstehen. Diese sind auch in den extrahierten Restsignalen hörbar. In der zweiten Hälfte des betrachteten Ausschnitts ist in beiden Fällen ein stimmhafter Klang wahrnehmbar (die Zeitachse ist relative zum Beginn des Ausatmens zu betrachten, startet also bei Sekunde 10 des Rohmaterials). Aufgrund der starken Behaftung mit Artefakten kann nicht mit Sicherheit festgestellt werden, welche Wortäußerung hier vorliegt. Als einzige objektive Methode zum Bestimmen der Aussage schlagen die Gutachter die Durchführung von Hörtests mit den separierten Audiosignalen vor. Das korrekte Design dieser Hörtests ist dabei ein wichtiger Punkt, z.B. sollten die Probanden möglichst kein Vorwissen über das Audiomaterial haben.

Der Fakt, dass das stimmhafte Signal so stark vom Ausatmen überdeckt wurde könnte darauf hinweisen, dass es in einiger Entfernung von den Mikrofonen ausgesprochen wurde und das Ausatmen zu diesem Zeitpunkt absichtlich zum Zweck der Verdeckung beigemischt wurde. Allerdings lassen sich im Signal keine Belege dafür finden.

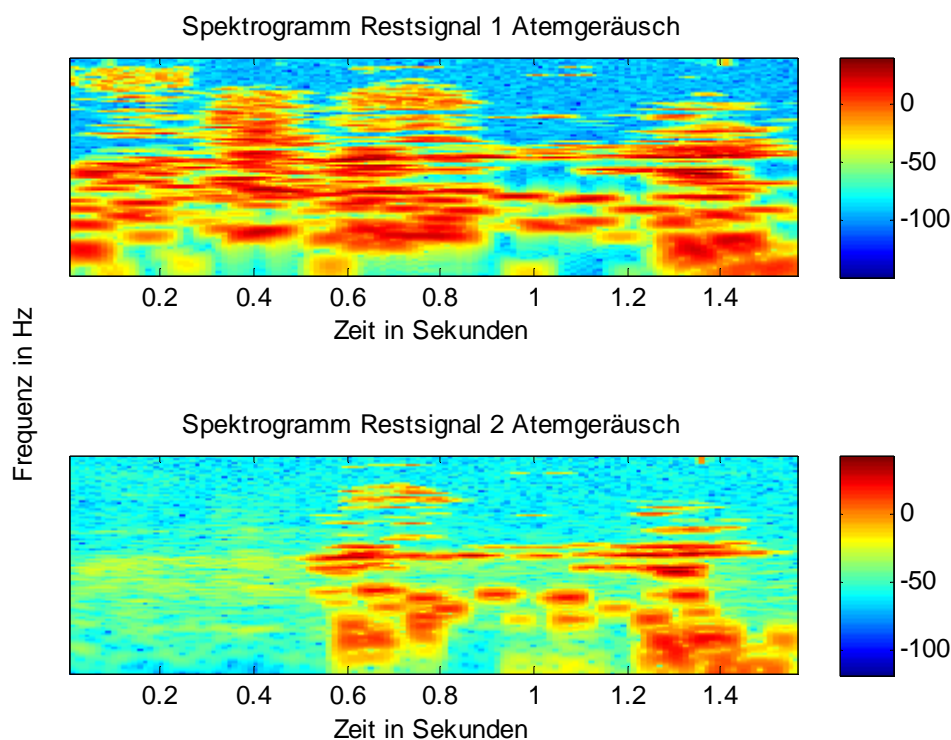


Abbildung 4: Spektrogramme der beiden Restsignale

⁶ Boll, S.F.: „Suppression of acoustic noise in speech using spectral subtraction“. IEEE Transactions on Acoustics, Speech and Signal Processing. Vol.. 27, No. 2, pp. 113-120, 1979.

5 Zusammenfassung

Die Gutachter kommen nach oben stehender Analyse zu dem Schluss, dass es sich bei dem vom ORF öffentlich zur Verfügung gestellten Rohmaterial um nachbearbeitetes Drehmaterial handelt. Hierfür spricht vor allem, dass sich die Tonmischungsverhältnisse während des Ausschnitts eindeutig ändern. Von unbearbeitetem Audiomaterial könnte man nur dann zweifelsfrei ausgehen, wenn alle Audiospuren der verschiedenen, an der Aufnahme beteiligten Mikrofone mit konstantem Mischungsverhältnis zusammengemischt werden, bzw. die einzelnen Spuren getrennt veröffentlicht werden.

Die Unterdrückung eines im Rohmaterial vorkommenden, auffälligen Ausatemgeräusches mittels spektraler Filterung offenbarte eine hintergründige Klangkomponente, der einem Stimmanteil ähnelt. Der Wortlaut kann jedoch abschließend nicht zweifelsfrei festgestellt werden. Hier kann nach Einschätzung der Gutachter nur ein Hörtest mit unabgängigen Probanden eindeutige Klarheit verschaffen.

6 Beschreibung der beigelegten Audiodaten

Die während des Gutachtens entstandenen Audiodateien sind in digitaler Form beigelegt. Ihr Inhalt und die Position innerhalb des Rohmaterials werden hier in Kurzform dokumentiert.

Dateiname	Beschreibung	Timecode (Ausschnitt)
rohmat_cropped_aligned.wav	Ausgeschnittenes Rohmaterial	00:00:00:00 bis 00:00:24:10
sendung_cropped_aligned.wav	Zum Rohmaterial synchronisierter Sendungsmitschnitt (nicht vorhandenes Material mit Stille aufgefüllt)	00:00:00:00 bis 00:00:24:10
separated_microphone_2.wav	Separierte Audiospur von Mikrofon 2	00:00:00:00 bis 00:00:07:00
filtered_breath_noise.wav	Restsignal von moderater Atemgeräuschunterdrückung	00:00:10:23 bis 00:00:11:83
strong_filtered_breath_noise.wav	Restsignal von starker Atemgeräuschunterdrückung	00:00:10:23 bis 00:00:11:83